



Health Information Technology Standards
Advisory Committee

Process for Implementing a Clinical-Grade Variant File for Reporting to Cancer Registries

Version 1.0

Use Case Prepared by:
Commonwealth of Virginia
Health Information Technology Standards
Advisory Committee (HITSAC)
Genomics Working Group

Developed in Partnership with
Centers for Disease Control and Prevention
Division of Laboratory Programs, Standards, and Services &
HL7 Clinical Genomics Working Group

September 23, 2014

Table of Contents

General Information.....	3
Clinical Grade Variant File.....	5
Implementing a Clinical-Grade Variant File for Reporting to Cancer Registries.....	6
Appendix A: Clinical-Grade Variant File Specification.....	X
Appendix B: <i>CAP Definition of Synoptic Reporting</i>	X
Appendix C: <i>HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model</i>	X
Appendix D: Clinical-Grade Variant File Work Flow.....	X
Appendix E: Clinical-Grade Variant File Data Flow.....	X
Appendix F: Data Elements for File Submission Log.....	X
Appendix G: Naming Conventions for File Log.....	X
Appendix H: Electronic Signature Requirements.....	X

General Information

1. Use Case ID

HITSAC Genomics Working Group Use Case #2 (HITSACGWG2)

2. Use Case Title

Process for Implementing a Clinical-Grade Variant File for Reporting to Cancer Registries

3. Abstract

The HITSAC Genomics Working Group Use Case #2 defines the process for implementing a clinical-grade variant file for the purpose of reporting genomic information to a central cancer registry (CR).

4. Description

The absence of an adopted standard for the specification, messaging and exchange of genomic information compromises or precludes a number of applications essential for integrating genomics into broader aspects of medical care. This gap in interoperability has implications at the clinical, laboratory and public health levels.

The Centers for Disease Control and Prevention (CDC) has formed a working group in conjunction with the College of American Pathologists (CAP), the HL7 Clinical Genomics Working Group and other partners to advance the uniformity in how genomic data generated from clinical Next-Generation Sequencing (NGS) tests are represented and made useful.

The CDC initiative has leveraged CAP's synoptic reporting specification as the basis for transmitting genomic information in a standardized manner. CDC also has reviewed how the CAP synoptic reporting specification aligns with HL7's LOINC-qualified genetic variation model. The CDC's intent is not to create a new data specification but to adapt what already exists into a clinical-grade variant file.

The following use case has been developed by the Commonwealth of Virginia's HITSAC Genomics Working Group (HITSAC-GWG) to define the process for implementing the CDC working group's clinical-grade variant file for the purpose of reporting genomic information to a central cancer registry (CR). The use case has been constrained to focus on reporting of machine-readable results from NGS platforms.

Future use cases identified by the HITSAC-GWG will concentrate on reporting results to CRs from tests conducted on other, non-NGS platforms. Future use cases also under consideration by the HITSAC-GWG will focus on implementing a clinical-grade variant file in human-understandable representations to enhance clinical utility and quality of patient care.

5. Stakeholders

- Cancer Registry (CR)
 - CR Software
 - CR Staff

- Certified Data Source (CDS)
 - CDS Software
 - CDS Staff

6. Definitions

Certified data source (CDS): A data source that the cancer registry has verified as able to perform electronic reporting by implementing the *NCPR-AERRO Certify a Data Source for Electronic Reporting Use Case*.

Clinical-grade variant file: A data file specification designed by the Centers for Disease Control and Prevention working group based on North American Association of Central Cancer Registries (NAACCR) standards, the College of American Pathologist's synoptic reporting specification and aligned with HL7's LOINC-qualified genetic variation model.

Event report: An electronic transmission of information to a cancer registry.

Record layout format: Describes how fields are positioned within a record.

7. Version Control

The following table contains a history of revisions to this publication.

Version	Date	Revision Description	Contact
1.0	9/23/2014	Initial Use Case Document	Joe Grubbs

Identifying Changes in This Document

- See the latest entry in the revision table above
- Vertical lines in the left margin indicate the paragraph has changes or additions. Specific changes in wording are noted using italics and underlines; with italics only indicating new/added language and italics that is underlined indicating language that has changed.

The following examples demonstrate how the reader may identify updates and changes:

Example with No Change – The text is the same. The text is the same.

Example with Revision – The text is the same. *A wording change, update or clarification is made in this text.*

Example of New Text – *This text is new.*

Clinical-Grade Variant File

1. Clinical-Grade Variant File Description

Note: The clinical-grade variant file applied in this use case has been included in this document as Appendix A.

The clinical-grade variant file developed by the CDC's working group XXXXX

[DR LUBIN PLEASE PROVIDE ADDITIONAL INFORMATION FOR THIS SECTION]

During clinical sequence analysis, patient specific sequence information is generated at two stages during the course of the test method. The first set of sequencing results is generated after the patient sample is sequenced, mapped and aligned against a human reference assembly, and analyzed for the presence of sequence variants. This initial set of sequence variants is analyzed to remove those of lower quality as defined by criteria set by the laboratory. This results in a set of high confidence sequence variants. This set is described within a variant file. This set of variants undergoes downstream analysis to identify the sequence variants that are clinically relevant for the patient with respect to the medical question asked that prompted the test order. Some laboratories will also analyze the initial data set of secondary findings that can have clinical relevance to the patient but typically not associated with the reason the test was ordered (Genet Med. 2013;15:566-574). Therefore, clinical sequencing test produce three data sets, 1) a set of high-confidence sequence variants, 2) variant(s) that are clinically relevant to the patient and the indication for testing, and 3) incidental findings.

For the purpose of this use case only variant(s) that are clinically relevant to the patient with respect to the indication for testing will be considered. This decision is predicated on the following:

1. Synoptic reports indicate variants and/or biomarkers relevant to the cancer under investigation, together with other relevant data. Structure does not exist to accommodate a broader range of genomic data although this can be something to explore. There are several types of synoptic reports in use and not all have been used or optimized for data extraction and electronic messaging.
2. Cancer registries consume data relevant to a defined set of medical conditions that directly correlate. With regards to genomic data, this would include the clinically relevant variants with respect to cancer diagnosis. The new paradigm is the capability to consume structured genomic data. Nonetheless, the process by which registries take in such data should be scalable to eventually allow the inclusion of high confident variant datasets or full exome/genome representation.
3. Current IT systems are generally not designed to message structured genomic data and metadata. Beginning with variants of clinical relevance with respect to the indication for testing provides the simplest model that can later be extended to host the high confidence variant set and incidental findings or a fuller exome/genome representation.

The data contained within variant files will form the basis for messaging of genomic data. Constraint of this data is required. The CDC facilitated Clinical-Grade Variant File Specification Word group is tasked to address this issue with respect to the broader objective of describing a variant file that complies with a minimal data standards sufficient to support interoperable systems. This will be accomplished in considering:

1. Data acquisition methods during clinical sequence testing prior to deposition of data into the variant file. This stage primarily addresses alignment strategies relevant to use of a standard coordinate system for describing variant and haplotype positions.

2. Data constraint and presence of metadata within the variant file. This stage addresses how certain data elements are described within the variant file.
3. Data translation to a format amenable for electronic messaging. This stage addresses data translation and representation consistent with established messaging standards, such as HL7. Specifically, this refers to such descriptors as HGNC and HGVS, for gene and variant representations. This also is the step in which LOINC codes would be assigned. The output from this translation is anticipated to reside in a file structure separate for the lab-generated variant file but it is conceivable that existing variant files can be enhanced to contain this additional data structure

The data to be messaged would comprise both sequence and metadata. The metadata is anticipated to contain sufficient information to permit the receiver to understand the sequence context in terms of

2. Alignment with Existing Health IT Standards

The clinical-grade variant file developed by the CDC working group and applied in this use case has been designed to align with the following health IT standards:

NAACCR Standards for Cancer Registries

Information on the North American Association of Central Cancer Registries (NAACCR) standards may be accessed on the organization's website at <http://naaccr.org/>.

CAP Synoptic Reporting Specifications

The CAP White Paper, *CAP Definition of Synoptic Reporting*, which defines synoptic reporting specifications for cancer data, has been provided in this document as Appendix B.

HL7 LOINC-Qualified Genetic Variation Model

The HL7 Implementation Guide, *HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model*, has been provided in this document as Appendix C.

Comment [IL1]: We should be inclusive of CAP synoptic reporting specifications but not to the extent to exclude others. This is because there are several types of synoptic reports and other means to store genomic data (in written format). I am told that no form of synoptic reporting has established itself as a standard for messaging genomic information. This said, we are working with CAP and the CDC Cancer Surveillance Branch to advance the use of CAP's synoptic reports for this purpose.

Comment [IL2]: As an FYI, LOINC coding for genetic test results have generally not been implemented in practice. A pathway to adoption would be an invaluable contribution

Implementing a Clinical-Grade Variant File for Reporting to Cancer Registries

Note: Diagrams of the work flow and data flow of the receive process for the clinical-grade variant file have been provided in this document as Appendix D and Appendix E.

1.0 Preconditions

A set of conditions that must be met before the activities described in the use case can **begin**.

1. The clinical-grade variant file is received electronically by a central cancer registry (CR).
2. The clinical-grade variant file entering the CR software is from a certified data source (CDS).

Comment [IL3]: As described above, the variant file noted here may be different that that which is generated during the course of clinical testing

2.0 Post Condition

A set of conditions that must be met after the activities described in the use case have been completed.

The received clinical-grade variant file has been accepted by the CR to go forward.

3.0 Priority

Describes the importance and sequence of the use case in the overall activities of the cancer registry.

This is a high-priority use case and the HITSAC Genomics Working Group decided it should be developed first.

4.0 Frequency of Use

Describes how often the activities in the use case take place.

The activities in this use case will take place each time a new or resubmitted clinical-grade variant file arrives at a CR from a CDS.

5.0 Normal Course of Events

Describes the specific steps taken to perform the activity in the use case.

Normal refers to the steps that are taken when everything goes according to routine procedures. Problems and exceptions are described in section 6, Alternate Course.

Business rules are statements that describe a decision that must be made and agreed to by those involved in the activity. In the context of this document, a business rule describes the decision that needs to be made, and in some circumstances provides a recommendation; in others, options for consideration and use.

Software requirements are statements that describe the functionality of the software that is required or recommended.

5.1 This use case begins when CR software retrieves the clinical-grade variant file.

5.2 CR software loads the clinical-grade variant file from a CDS. [BR01]

BR	Business Rule	Purpose	Remarks
01	The time interval for processing a clinical-grade variant file from a CDS should be within five (5) days.	To ensure timely reporting.	This use case can be performed at a different interval than subsequent use cases. The time interval must be very short to identify and resolve problems before several clinical-grade variant files have been submitted with the same problems.

Comment [IL4]: Need to check is there are situations requiring a more immediate result (e.g., newborn screening)

5.3 CR software decrypts the clinical-grade variant file.

Note: This use case assumes that all clinical-grade variant files will be encrypted to meet Commonwealth of Virginia standards to ensure patient privacy, security and confidentiality.

5.4 CR software logs the clinical-grade variant file as received. [BR02, BR03]

BR	Business Rule	Purpose	Remarks
02	The clinical-grade variant file log should include recommended data items.	To ensure the ability to track and monitor file submissions accurately.	See Appendix F for a list of data items to include in the file log.
03	A standard naming convention should be used for the clinical-grade variant files.	To provide a national naming standard and track the files submitted.	See Appendix G for the proposed format.

Comment [IL5]: Just a reminder note to indicate that certain conventions are described by HL7

5.5 CR software stores the clinical-grade variant file in a temporary workspace on the CR computer system.

5.6 CR software validates the record layout format for the clinical-grade variant file. [BR04, BR05, BR06]

BR	Business Rule	Purpose	Remarks
04	The CDS must submit event reports using NAACCR or another nationally recognized standard messaging / record layout format.	To achieve uniformity and consistency.	<p>Cancer registry abstracts: The appropriate edition and version of the <i>NAACCR Standards for Cancer Registries</i></p> <p>Pathology laboratories: The appropriate edition and version of the <i>NAACCR Standards for Cancer Registries</i></p> <p>Billing and claims data: Uniform Billing Standard ANSI ASC X12</p>

5.7 CR software determines that the clinical-grade variant file is not a duplicate of a previous submission. [BR05]

BR	Business Rule	Purpose	Remarks
05	An electronic signature for the clinical-grade variant file as a whole should be created and stored in the database. See Appendix H.	To prevent reprocessing of clinical-grade variant file submissions.	<p>The electronic signature prevents a clinical-grade variant file from being processed more than once. Scenarios include:</p> <ul style="list-style-type: none"> ▪ A CDS may submit the same file multiple times. ▪ The CR may mistakenly try to process the same file twice.

5.8 CR software determines there are no exact duplicate event reports in the clinical-grade variant file submission. [BR06]

BR	Business Rule	Purpose	Remarks
06	<p>CR software should perform a deterministic record-by-record and data item-by-data item match.</p> <p>There may be a performance issue to check pathology reports data item-by-data item, so a subset of data items may be used.</p>	To confirm that the clinical-grade variant file is a new submission.	<p>Subset of data items.</p> <p>Same reporting source:</p> <ul style="list-style-type: none"> ▪ Last name ▪ First name ▪ Sex ▪ Date of birth ▪ Primary site ▪ Laterality ▪ Date of diagnosis ▪ Morphology (histology/behavior)

Comment [IL6]: The same histology sample may be used for multiple tests. Therefore, we need to carefully investigate what this subset would look like as not to exclude new test results

5.9 CR software loads the clinical-grade variant file into the CR database, and the use case ends.

6.0 Alternate Course of Events

Numbering in this section refers to its associated step above in section 5, Normal Course of Events.

5.6a The clinical-grade variant file is not in a valid record layout format.

- 5.6a.1 CR software rejects the clinical-grade variant file.
- 5.6a.2 CR software notifies the CDS (software) that clinical-grade variant file is rejected.
- 5.6a.3 CR software records reason for rejection and updates clinical-grade variant file log.
- 5.6a.4 End of use case.

5.7a The clinical-grade variant file is an exact duplicate of a previously submitted clinical-grade variant file. [BR05]

- 5.7a.1 CR software marks it as a duplicate.
- 5.7a.2 CR software rejects the clinical-grade variant file.
- 5.7a.3 CR software notifies the CDS (software) that clinical-grade variant file is rejected.
- 5.7a.4 CR software records reason for rejection and updates clinical-grade variant file log.

BR	Business Rule	Purpose	Remarks
05	An electronic signature for the clinical-grade variant file as a whole should be created and stored in the database. See Appendix H.	To prevent reprocessing of clinical-grade variant file submissions.	The electronic signature prevents a clinical-grade variant file from being processed more than once. Scenarios include: <ul style="list-style-type: none"> ▪ A CDS may submit the same file multiple times. ▪ The CR may mistakenly try to process the same file twice.

5.8a The clinical-grade variant file contains exact duplicate event reports. [BR06]

- 5.8a.1 CR software marks the event report as a duplicate, inserts in the duplicates table and deletes it from the clinical-grade variant file.
- 5.8a.2 End of use case.

BR	Business Rule	Purpose	Remarks
06	CR software should perform a deterministic record-by-record and data item-by-data item match. There may be a performance issue to check pathology reports data item-by-data item, so a subset of data items may be used.	To confirm that the clinical-grade variant file is a new submission.	Subset of data items. Same reporting source: <ul style="list-style-type: none"> ▪ Last name ▪ First name ▪ Sex ▪ Date of birth ▪ Primary site ▪ Laterality ▪ Date of diagnosis ▪ Morphology (histology/behavior)

Comment [IL7]: The same histology sample may be used for multiple tests. Therefore, we need to carefully investigate what this subset would look like as not to exclude new test results

7.0 Business Rules

A statement that describes a decision that must be made and agreed to by those involved in the activity. In the context of this document, a business rule describes the decision that needs to be made, and in some circumstances provides a recommendation; in others, options for consideration and use.

Business rules for this use case are presented under the step to which they apply.

BR	Business Rule	Purpose	Remarks
01	The time interval for processing a clinical-grade variant file from a CDS should be within five (5) days.	To ensure timely reporting.	This use case can be performed at a different interval than subsequent use cases. The time interval must be very short to identify and resolve problems before several clinical-grade variant files have been submitted with the same problems.
02	The clinical-grade variant file log should include recommended data items.	To ensure the ability to track and monitor file submissions accurately.	See Appendix F for a list of data items to include in the file log.
03	A standard naming convention should be used for the clinical-grade variant files.	To provide a national naming standard and track the files submitted.	See Appendix G for the proposed format.
04	The CDS must submit event reports using NAACCR or another nationally recognized standard messaging / record layout format.	To achieve uniformity and consistency.	Cancer registry abstracts: The appropriate edition and version of the <i>NAACCR Standards for Cancer Registries</i> Pathology laboratories: The appropriate edition and version of the <i>NAACCR Standards for Cancer Registries</i> Billing and claims data: Uniform Billing Standard ANSI ASC X12
05	An electronic signature for the clinical-grade variant file as a whole should be created and stored in the database. See Appendix H.	To prevent reprocessing of clinical-grade variant file submissions.	The electronic signature prevents a clinical-grade variant file from being processed more than once. Scenarios include: <ul style="list-style-type: none"> A CDS may submit the same file multiple times. The CR may mistakenly try to process the same file twice.
06	CR software should perform a deterministic record-by-record and data item-by-data item match. There may be a performance issue to check pathology reports data item-by-data item, so a subset of data items may be used.	To confirm that the clinical-grade variant file is a new submission.	Subset of data items. Same reporting source: <ul style="list-style-type: none"> Last name First name Sex Date of birth Primary site Laterality Date of diagnosis Morphology (histology/behavior)

Comment [1L8]: 5 days may be too long for some tests (e.g. newborn screening?).

8.0 Exceptions

None.

9.0 Includes

None.

10.0 Special Requirements

None.

11.0 Assumptions

1. The CDC working group has completed specification of the clinical-grade variant file.
2. Clinical-grade variant files are in electronic format.
3. Clinical-grade variant files are encrypted pursuant to Commonwealth of Virginia standards.

12.0 Pilot Project Recommendation

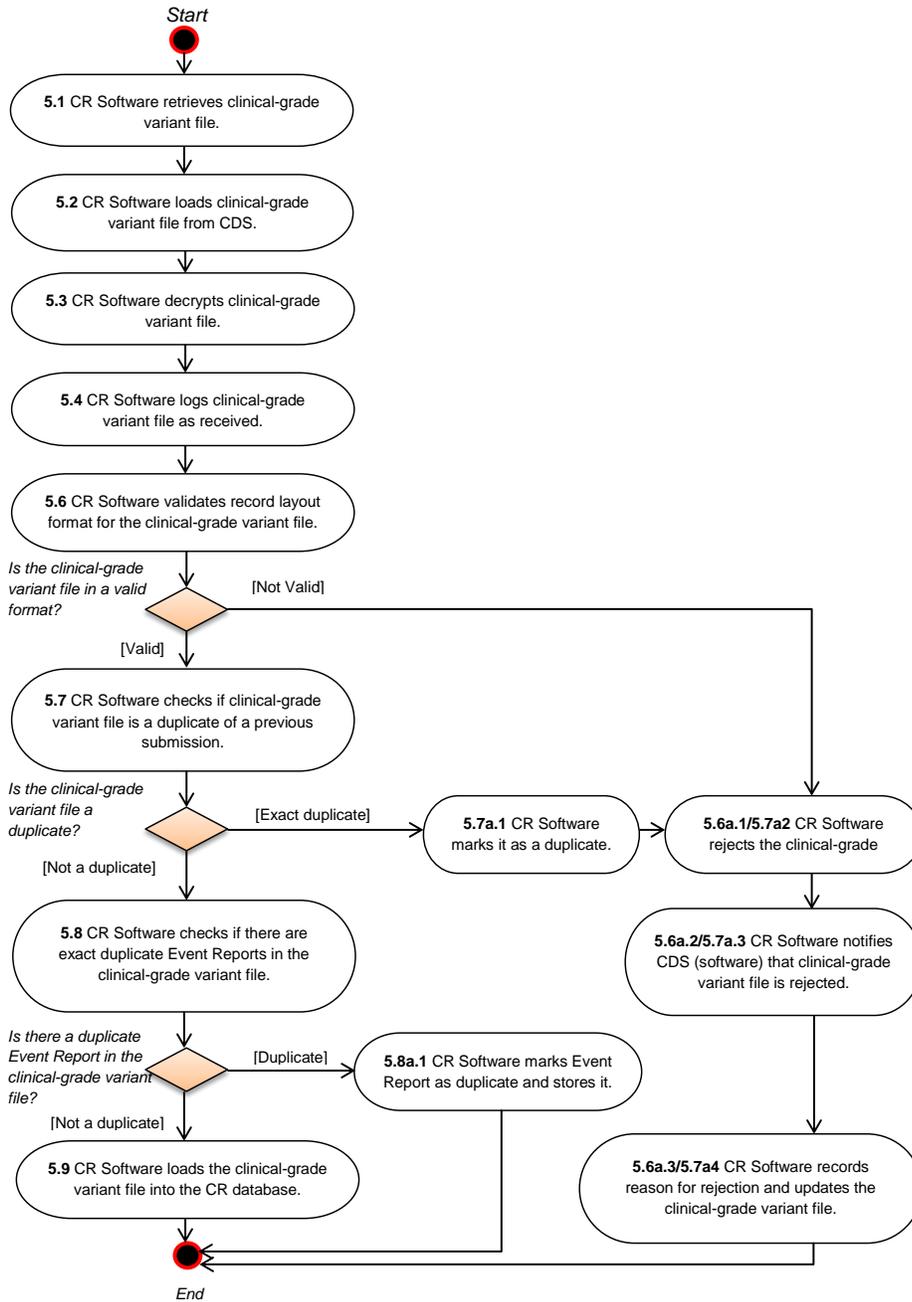
The findings from this use case will be used to frame a pilot project to test the process for implementing a clinical-grade variant file for the purpose of reporting genomic information to a central CR. The proposed pilot project has been documented in the white paper, *HITSAC Genomics Working Group Pilot Project Recommendation: Process for Implementing a Clinical-Grade Variant File for Reporting to Cancer Registries*. The purpose of the pilot project will be to develop and implement a data model and control log to support the clinical-grade variant file submissions and an implementation guide to document the process for using the file for CR reporting.

13.0 References

1. Baseline use case content and the clinical-grade variant file specification provided by the Centers for Disease Control and Prevention Division of Laboratory Programs, Standards, and Services.
2. *HL7 Domain Analysis Model: Clinical Genomics, Release 1*. HL7 Informative Ballot, September 2014.
3. *HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model, Release 1*. HL7 Informative Ballot, April 2009. **[CITATION FOR FINAL VERSION]**
4. *NAACCR Standards for Cancer Registries*, Applicable Volumes.

NOTE: FULL SET OF APPENDICES WILL BE ADDED IN PRODUCTION VERSION

Appendix D: Receive Clinical-Grade Variant File Workflow Diagram



Appendix E: Receive Clinical-Grade Variant File Data Flow Diagram

